DOCUMENT RESUME

ED 282 893 TM 870 266

AUTHOR Terwilliger, James S.
TITLE Classroom Evaluation Practices of Secondary Teachers in England and Minnesota.
PUB DATE Apr 87
NOTE 28p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Rating; Cross Cultural Studies; Educational Practices; *Evaluation Criteria; *Evaluation Methods; Foreign Countries; *Grading; High School Students; Mathematical Models; School Surveys; Secondary Education; *Secondary School Teachers; *Student Evaluation
IDENTIFIERS England; Minnesota; *Survey of Grading Practices

ABSTRACT

This study was designed to obtain descriptive data on the classroom evaluation practices of secondary teachers in two different public school systems. A self-report survey instrument was employed to collect data from 217 teachers in York, England, and 217 teachers in Minnesota. Respondents were asked to indicate which specific techniques they employed and to rate the relative importance of a variety of factors in judging achievement for the purpose of assigning subject matter grades. Analyses of the responses according to four log linear models reveal important interactions among teachers grouped by subject matter (English, Foreign Language, Mathematics, Science, and Social Studies) and by country. The survey results indicate: (1) teachers' judgments were based on a variety of quasi-measurement techniques, as well as on tests; (2) ratings of importance of various types of data reveal that both the educational system and the subject matter have an important bearing upon the relative weight given to specific information; (3) ratings of the remaining factors employed in grade assignment are jointly influenced by both the educational system and the subject matter field; and (4) the primary basis selected by teachers for the assignment of grades is itself significantly influenced by differences in both educational system and subject matter taught. The survey instrument is appended. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED282893

Classroom Evaluation Practices of Secondary Teachers in England and Minnesota

James S. Terwilliger

University of Minnesota

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. S. Terwilliger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at NCME Annual Meetings, Washington, D.C., April 23, 1987.

TM 870 266

Classroom Evaluation Practices of Secondary Teachers in England and Minnesota

Abstract

This study was designed to obtain descriptive data on the classroom evaluation practices of secondary teachers in two quite different public school systems. A self-report survey instrument was employed to collect data from 217 teachers in England and 217 teachers in Minnesota. Respondents were asked to: 1) indicate which specific techniques they employ, and 2) rate the relative importance of a variety of factors in "judging achievement for the purpose of assigning subject matter grades." Analyses of the responses reveal important interactions among teachers grouped by subject matter (English, Foreign Language, Mathematics, Science, and Social Studies) and by country.

Introduction

In a recent review of research on classroom assessment, Stiggins, et. al. (1986) note that the great preponderance of published research studies focus upon the use of standardized tests to the exclusion of teacher-made assessment devices. This is true despite the fact that classroom teachers regularly design, administer, and employ results from a wide variety of assessment devices in making important decisions about students they teach. Ironically, a special 1983 issue of the Journal of Educational Measurement (Burnstein, 1983) on the state of art in linking testing and instruction explicitly excluded any consideration of teacher-made tests.

Previous Studies

What little is known about the day-to-day assessment practices of classroom teachers is the result of survey studies reported by Gullikson (1984), Gulllikson and Ellwein (1985), and Stiggins and Bridgeford (1985). The Gullikson studies deal specifically with issues related to the use of teacher-made tests based upon responses of a random sample of over 300 South Dakota teachers. A primary objective of the research was to learn more about how teachers perceive and use tests with a view to improving teacher training programs in measurement.

Stiggins and Bridgeford (1985) address somewhat more general issues associated with the use of a broader variety of assessment practices in the classroom: a) teacher-made objective tests, b) published tests, and c) observations and ratings of student behavior and products. Responses were obtained from over 200 teachers from eight different school districts located in geographically dispersed regions of the U.S. As with the Gullikson articles, a primary goal of the research was to identify practices and stated concerns of teachers which have direct implications for improving the training of teachers in classroom assessment.

The Current Study

This study reports a survey of classroom assessment practices related specifically to one type of decision commonly made by all classroom teachers -- the assignment of grades to students. The survey focuses upon grading of students because the assignment of grades is almost universally mandated by schools and grades play an important role in the lives of both students and teachers. In particular, the project deals specifically with classroom evaluation practices of secondary teachers in the five traditional "core" subject matters: English, Foreign Language, Mathematics, Science, and Social Studies.

The study is unique in that the same survey instrument was employed with teachers in two very different settings, each with it's own distinctive educational history and philosophy. Responses were obtained in Comprehensive Secondary Schools in England and a second sample of over 200 teachers in public secondary Schools in Minnesota.

The following questions are addressed by this study:

1. What assessment techniques are most commonly employed by teachers for the purpose of grade assignment?
2. What relative weights do teachers attach to various types of data in assigning grades to students?
3. What explicit or implicit frames of reference do teachers employ when assigning grades to students?
4. What differences in assessment practices exist among teachers as a function of:
   a. subject matter field?
   b. type of educational system (England vs Minnesota)?
5. What, if any, interactions exist among subject matter field and educational system as factors influencing:
   a. relative weights attached to various types of data in grade assignment?
   b. frames of reference employed in assigning grades to students?

Data Collection Procedures

Survey Form

A copy of the survey form appears in Appendix A. The form was designed with a separate response sheet on which participants recorded answers to all questions. No information which would identify individual respondents was requested. The survey form was designed to require no more than 30 minutes to complete in an effort to gain maximum cooperation from respondents. Consequently, the number and type of responses requested were restricted to issues associated with classroom assessment for purposes of grade assignment.

Data collection was accomplished by mailing packets of survey forms and response sheets to schools that had authorized teachers to participate in the survey. Materials were distributed to teachers in each school by a "coordinator" designated by the school principal (or Headmaster). Completed response sheets were returned by the "coordinators" in pre-stamped self-addressed envelopes.

Survey Respondents

English Sample

Data were collected in England while the author was a Visiting Fellow in the Department of Education at the University of York in the fall of 1983. Prospective survey respondents were contacted through their schools. Schools were selected in two different ways. First, a list of 15 comprehensive secondary schools was identified with the aid of the staff of the Centre for the Study of Comprehensive Schools (CSCS) at the University of York.[1,2] These schools were thought to be especially likely to cooperate in the survey because of previous involvement with

---

[1]The author is grateful to Frank Stoner and Ben Kerwood of the Centre for the Study of Comprehensive Schools for providing invaluable help in identifying and contacting schools for participation in the project.

[2]Comprpehensive schools in England are publicly supported co-educational secondary schools serving the needs of the general community in a fashion similar to public secondary schools in the United States.

activities sponsored by CSCS.

A second group of 12 comprehensive secondary schools was selected at random. These schools were included to create a more scientifically acceptable sample and to check the possibility that the schools identified by the CSCS staff might yield unrepresentative results.

Letters requesting cooperation in the survey were sent to both the Chief Education Officer of the appropriate education authority and the Head of each school. Nine of the 15 schools from the CSCS list and seven of the 12 schools from the random sample agreed to participate in the study. Participation at the level of the individual school was strictly voluntary. The number of responses received from an individual schools ranged from five to 37. A total of 217 completed response sheets was received from the 16 participating schools.

Minnesota Sample

Prospective survey respondents were contacted through their schools in the fal of 1984. First, a list of nine school systems in the Twin Cities region was identified. These nine school systems were thought to be a representative cross section of school systems within the Twin Cities metropolitan area in terms of size and geographical location.

A phone call was made to the administrative office of each school system in order to identify the appropriate individual to contact concerning a request for cooperation in the survey. A letter describing the survey in more detail along with a draft copy of the survey form was sent to the contact person in each school system. Five school systems subsequently agreed to participate.

In each of the participating systems, the names of specific school principals were provided. The school principals were next contacted in order to identify a survey "coordinator" within each school building. A total of 21 schools participated in the survey. Individual teachers participated on a strictly voluntary

basis. The number of completed response forms received from individual schools ranged from three to 26. The total number of responses received was 217.

A cautionary note is in order before turning to the results. The fact that both the schools and individual teachers participated on a strictly voluntary basis may introduce a degree of bias into the results. The nature and extent of such bias is likely to vary from one specific question to the next. Hopefully, the error due to sampling bias is not large and the magnitude of the error is comparable in the two major samples.

It should also be noted that all responses given were with reference to a specific class taught by the respondent. Teachers were asked to name "the first course that you teach on Mondays" as a means of arbitratily designating a classroom context for responding to the survey. The course designated was the basis for subsequent classification of teachers according to subject matter fields. Consequently, it is impossible to say to what extent the assessment practices of a given respondent may vary from one instructional context to another.

Results

Comparison of Two English Samples

The first step in the analysis of the data was to compare the responses of teachers from the nine English schools identified by the CSCS staff with those of teachers from the seven English schools selected at random. If the two groups differ in ways that are likely to affect the interpretation of the results, separate analyses should be performed on each. However, if there are no importanat differences between the two groups, analysis of the combined data for all respondents is appropriate.

The responses of the two groups were compared on each of 29 items in the survey form dealing with assessment practices. Differences ($p < .05$) were found on only five items. Comparisons were also made using biographical information about

respondents (sex, age, educational background) and courses being taught by respondents (subject area, year level, pupil ability level, etc.) in the two groups. Although there were differences on certain of these variables the differences appear to be unrelated to differences in classroom assessment practices. Consequently, the two groups were combined into a single "English" sample for purposes of subsequent analyses. The results of these analyses are reported in the following sections.

Biographical Data

Table 1 presents a breakdown of the English and Minnesota samples by sex, age, years of teaching experience, and educational background. A significantly higher proportion of the Minnesota teachers are female ($p < .01$) and the Minnesota teachers are significantly older ($p < .001$) and more experienced ($p < .001$). Generally speaking, the educational level of the Minnesota teachers is also higher but differences in the post-secondary educational systems in England and the United States make it difficult to make clean comparisons on this variable.

Insert Table 1 about here

Evaluation Techniques Employed

The first section of the survey form presented the respondents with a list of 16 specific evaluation techniques (grouped under four major headings) and requested that they indicate each technique used in their own classrooms (see Appendix A). Table 2 gives a summary of responses when teachers were asked to check all classroom data collection techniques employed. A higher proportion of the Minnesota sample reports the use of 12 of the 16 techniques listed. For eight of these the differences are 10% or greater. Differences of this magnitude are statistically significant ($p < .01$). For the four techniques which are used with higher frequency by the English sample, the differences are all less than 10%. None is statistically significant.

It comes as no surprise that two general categories of techniques are most commonly employed in both samples: a) written tests and examinations, and b) records of assigned work, classroom participation, and major projects. Within the first category, short answer/completion questions are most popular. The Minnesota teachers indicate a stronger preference for multiple choice/true-false and matching questions than do the English sample. Within the second category, records of assigned work (both classwork and homework) are employed with the highest frequency in both samples. But Minnesota teachers use records of assigned work, classroom participation, and major out-of-class projects with higher frequency than do teachers in England.

Insert Table 2 about here

Relative Importance of Factors in Grade Assignment

The second section of the survey form includes a list of 12 different types of information and/or factors frequently employed by teachers when judging student achievement for the purpose of assigning grades (see Appendix A). Respondents were instructed to, "Indicate how much consideration you give to each factor by selecting one of four response categories: NONE (0%); MINOR (1-10%); MODERATE (11-25%); and MAJOR (26-100%)."

The resulting data can be cast into a three-dimensional table based upon a cross-classification by educational system (England vs. Minnesota), subject matter (English, foreign language, mathematics, science, and social studies), and ratings of importance (none, minor, moderate, major). An illustration of such a table is presented in Table 3 where the factor rated is "Written Tests."

As previously stated, a primary question of interest is, "What interactions exist among subject matter field and educational system as factors influencing relative weights attached to various types of data in grade assignment?" To answer

this question a series of log-linear analyses were performed treating educational system and subject matter field as independent variables and rating of importance as the dependent variable. In particular, the analysis employed an hierarchical approach (Knoke & Burke, 1980) starting with the simplest model and successively increasing the model complexity by adding more terms until an acceptable "fit" is obtained as defined by a χ^2 statistic. This will be illustrated with reference to Table 3.

Insert Table 3 about here

Since there are 40 cells (2x5x4) in the three-way breakdown, there are 40 - 1 = 39 total degrees of freedom available in the log-linear analysis. Any proposed model for estimating expected cell frequencies employs a set of parameters with one degree of freedom associated with each parameter used. The χ^2 statistic resulting from the comparison of the observed and expected cell frequencies has an associated degrees of freedom equal to 39 minus the number of parameters employed in generating the expected frequencies.

Generally speaking, the simplest model is the one which employs the fewest parameters in estimating the cell frequencies. Such a model would imply no relationship between the dependent variable and the independent variables. For the data in Table 3 the simplest model (Model 1) is one which postulates that there are no interactions involving either ratings of importance and educational system (England vs. Minnesota) or ratings of importance and subject matter field (English, foreign language, mathematics, science and social studies). This means that the model consists of "main effects" due to each of the three dimensions plus "interactions" due to various combinations of educational system and subject matter field. Specifically, the sources and number of the parameters employed in Model 1 are summarized as follows:

Model 1: (System, Subject) (Rating)

Source	Parameters	df
Educational System	1	1
Subject Matter	4	4
System X Subject Matter	4	4
Importance Rating	3	3
Model Total	12	12

Notice that the number of parameters employed for each main effect is the number of levels on that dimension minus one. As with ANOA models, the number of parameters associated with interactions is the product of the number of parameters used in each of the two relevant main effects. The df associated with the χ^2 statistic for Model 1 is 39 - 12 = 27.

The χ^2 which resulted from employing Model 1 with the data in Table 3 is 68.88. A value as large as this is highly significant ($p < .001$) with 27 df. Consequently, the logical conclusion is that the model does not fit. Hence, we reject the hypothesis that there are no interactions involving either ratings of importance of written tests and educational system or ratings of importance of written tests and subject matter field. In other words, it must be true that one (or both) of the independent variables has(have) an influence upon the dependent variable. The nature of the influence can be seen by examination of Table 3. Note that Minnesota teachers rate written tests as more important than do teachers in England. Also, foreign language, math, and science teachers in both samples tend to rate written tests as more important than do teachers of English and social studies. The magnitude of these influences can be determined by examining the fit of more complex models.

A slightly more complex model can be created by assuming that the educational system is an important influence upon ratings of the importance of written tests. This hypothesis is supported by the fact that Minnesota teachers tend to rate

written tests as more important than do the teachers from England (see Table 3). This model adds three interaction parameters (1 (system) X 3 (rating)) to the previous model and is summarized as:

Model 2: (System, Subject) (System, Rating)

Total Parameters = 15

df for χ^2 = 24

The χ^2 associated with Model 2 is 50.87. This is highly significant ($p < .005$) for 24 df suggesting that the model provides a poor fit to the observed results. Formally, we reject the model as an explanation of the data and conclude that factors other than educational system must influence ratings of the importance of tests.

An alternative to Model 2 is a model which incorporates parameters associated with subject matter field (rather than educational system) as an influence upon ratings of the importance of written tests. Such a model adds twelve interaction parameters (4 (subject) X 3 (rating)) to Model 1. This is:

Model 3: (System, Subject) (Subject, Rating)

Total Parameters = 24

df for χ^2 = 15

The χ^2 for Model 3 is 26.49. This is clearly a better fit than the previous models as evidenced by the drop in the χ^2 statistic. However, the obtained χ^2 is statistically significant ($p < .05$) in terms of the 15 df associated with it. Once again, we reject the hypothesis that the model provides an adequate "fit" for the observed data and consider an even more complex model.

The next most complex model is one which combines all the parameters employed in the three previous models. This model is premised on the assumption that <u>both</u>

educational system and subject matter play an important role in teacher ratings of the importance of written tests. In brief, the model includes parameters for the three main effects and all the two-way interactions among the three dimensions of Table 3. The model is summarized as:

Model 4: (System, Subject) (System, Rating) (Subject, Rating)

Total Parameters = 27

df for χ^2 = 12

The χ^2 for Model 4 is 9.63. This value is dramatically lower than the previous χ^2-values. The p-value associated with this χ^2 is greater than .25. Consequently, we conclude that there is a reasonable "fit" between the expected frequencies generated by the model and the observed frequencies in Table 3.

The foregoing series of analyses lead to the conclusion that teacher ratings of the importance of written tests in grade assignment are a function of both the subject matter field of the teacher and the educational system in which the teacher teaches.

A series of analyses using the four log-linear models described above was performed on each of the remaining eleven factors rated for importance. The results of these analyses are summarized in Table 4. Chi-square values accompanied by an asterisk have p-values between .05 and .10, indicating a marginal or "weak fit" to the data. Chi-square values accompanied by a double asterisk have p-values greater than .10 indicating a "better" fit to the data.[3] Values in parentheses are for higher order models in cases where a simpler model provides a reasonable fit. Nor-

Insert Table 4 about here

---

[3] Although there are no universally accepted rules for deciding when a model "fits", Knoke and Burke (1980) suggest that models with p-values "between .10 and .35" be accepted as fitting the data.

mally, these would not be of interest. They are reported here strictly for comparison purposes.

It is clear from the values in the first column of Table 4 that Model 1 did not yield a fit for any of the factors rated. In other words, all ratings of importance are influenced by one or both of the independent variables.

Examination of column two of Table 4 reveals that the educational system (England vs. Minnesota) has a substantial influence upon the ratings of six factors. In particular, the last four factors listed are rated very differently by teachers in England and those in Minnesota with ratings being much higher for teachers in England. Model 2 results in a fit to the data ($p > .10$) for these factors because it includes parameters relating educational system to ratings. Model 2 also results in a "weak" fit for ratings of assigned homework which was also rated as more important by teachers in England. Despite the fact that a fit is not achieved, ratings of end-of-term exams are greatly influenced by the educational system as can be seen by the substantial drop in the χ^2-value from Model 1 to Model 2. Again, teachers in England give this factor a much higher rating of importance.

The influence of subject matter field upon ratings of importance can be seen in column three of Table 4. Here there is a "weak" fit for ratings of assigned homework. This results from the fact that science teachers tend to rate homework as less important than do teachers in other subjects. The fit of the ratings of oral exams/class presentations is quite good since the χ^2-value has $p < .10$. The fit in this case is due to the fact that teachers of foreign language and English rate this factor as more important than do teachers in the other subject matter fields.

Ratings of four other factors reveal that there are subject matter differences although Model 3 fails to provide a fit to the data. These are seen in substantial drops in the χ^2-value from Model 1 to Model 3. The factors are: assigned classwork, class participation, major projects, and practical demonstrations.

Factors which require inclusion of both independent variables (Model 4) to achieve a fit to the observed ratings are: end-of-term exams (weak fit), assigned classwork, assigned homework, class participation, and major projects. In all cases, inspection of the three-way contingency tables reveals differences both between the two samples and among the subject matter groupings within each sample. One factor, practical demonstrations, did not achieve any fit for Model 4. This suggests that an even more complex model incorporating three-way interactions would be necessary in this case. No attempt was made to fit such a model.

Basis for Assigning Grades

The final section of the survey consisted of a single question designed to determine the fundamental viewpoint teachers adopt when assigning grades reflecting student achievement (see Appendix A). Specifically, the questions asked respondents to indicate which one of four alternative bases "best represents the primary basis you use" in assigning grades. The alternatives given were:

1. achievement with respect to individual student ability,
2. achievement with respect to other students of comparable age and grade level,
3. achievement with respect to absolute standards of performance, and
4. achievement with respect to individual student improvement or growth.

The same series of hierarchical log-linear analyses previously described were applied to responses to this section of the survey. The results are: Model 1, $\chi^2 = 92$ ($p < .001$); Model 2, $\chi^2 = 55$ ($p < .001$); Model 3, $\chi^2 = 53$ ($p < .001$) and Model 4 $\chi^2 = 18$ ($p > .10$). Therefore, the choice of the primary basis chosen by a teacher for grade assignment is influenced by the two independent variables in combination while neither one alone provides a reasonable fit to the data. Inspection of responses reveals that a higher proportion of Minnesota teachers (38%) prefer "absolute standards" than do teachers in England (12%) while the reverse is true for grading based upon "individual student ability" (12% vs. 19%), "other students" (41%

vs. 48%) and "growth" (9% vs. 20%). In both samples science teachers demonstrate a strong preference (60% - 70%) for basing grades upon "other students" while teachers of social studies show great variation with no one basis being preferred by more than 35% of that group. Over 60% of the teachers of foreign language and mathematics in the English sample indicate a preference for "other students" as a basis for grade assignment. Conversely, teachers of foreign language and mathematics in the Minnesota sample have a definite preference (over 50%) for "absolute standards" when assigning grades.

Implications of Results

The survey results clearly indicate that secondary teachers employ a wide variety of data in making judgments about student grades. Some of the most common sources of data (classwork, homework, class participation and major projects) depend upon quasi-measurement techniques which are largely ignored in the measurement literature and are given scant attention in most textbooks on classroom assessment. This suggests a need for a more systematic treatment of these sources of data in an effort to enhance the reliability and validity of the judgments which result. On the other hand, the strong preference of Minnesota teachers for "objective type" tests and examinations (short answer, multiple-choice, true-false and matching) probably reflects the attention that these item formats receive in American textbooks on classroom measurement.

Ratings of importance of various types of data reveal that both the educational system and the subject matter have an important bearing upon the relative weight given to specific information. A set of "non-cognitive" variables (class attendance, student "effort", classroom behavior, and work habits) are clearly regarded as more important in grading by teachers in England than teachers in Minnesota. It is likely that this reflects basic differences in the traditions and values of the two respective school systems. On the other hand, ratings of oral examinations and

class presentations appear to be a function of differences among subject matter fields rather than educational systems. This is not surprising given the traditional differences in the performance criteria in language classes as opposed to other subjects.

Ratings of the remaining factors employed in grade assignment (written tests, end-of-term exams, classwork, homework, class participation and major projects) are jointly influenced by both the educational system and the subject matter field. Consequently, classroom grading practices which depend upon these six factors are a function of both variables. The complexity which these results suggest is typically ignored by authors of standard textbooks on classroom assessment.

Finally, the primary basis selected by teachers for the assignment of grades is itself significantly influenced by differences in both educational system and subject matter taught. Subject matter differences, in particular, have importance for scholars concerned with applied educational measurement issues. There is currently little consensus across subject matters as to the most desirable basis for assigning grades. In fact, science is the only subject matter in which there is a high degree of agreement ("other students") across both samples. Since the basis teachers select for grade assignment frequently reflects a fundamental philosophical position concerning "what is fair" to students, we need to learn much more about the origins of such beliefs before attempting to change teachers' practices.

References

Burnstein, L. (1983). A word about this issue. Journal of Educational Measurement, 20(2), 99-101.

Gullikson, A. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77(4), 244-246.

Gullikson, A. and Ellwein, M. (1985). Post-hoc analysis of teacher-mode tests: The goodness-of-fit between prescription and practice. Educational Measurement: Issues and Practice, 4(1), 15-18.

Knoke, D. and Burke, P. (1980). Log-linear models. Quantitative Applications in the Social Sciences, #20, Sage Publications, Inc., Beverly Hills, CA.

Stiggins, R. and Bridgeford, N. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.

Stiggins, R., Conklin, N. and Bridgeford, N. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice, 5(2), 5-17.

Table 1

Summary of Biographical Data for Survey Respondents[a]

Sex	MN N	MN %	ENGLAND N	ENGLAND %
Female	93	43	67	31
Male	123	57	150	69
	216	100	217	100

Age	MN N	MN %	ENGLAND N	ENGLAND %
21-30	15	7	47	22
31-40	54	25	106	49
41-50	78	36	45	21
Over 50	68	32	17	8
	215	100	215	100

Years of Teaching	MN N	MN %	ENGLAND N	ENGLAND %
1 or 2	7	3	12	6
3-5	9	4	24	1
6-10	16	7	69	32
Over 10	185	85	110	51
	217	100	215	100

Educational Background	MN N	MN %	ENGLAND N	ENGLAND %
BA (or equiv)	75	35	13	52
Masters (or above)	131	65	103	48
	212	100	216	100

[a] Because some respondents left certain questions blank, the total N varies from item to item. The percents given are based upon those responding to each question and therefore always sum to 100.

Table 2

Summary of Evaluation Techniques
Employed by Survey Respondents

	%[a] MN	%[a] ENGLAND
A. Written tests and examinations		
1. Essay questions	51	53
2. Short answer and/or completion (fill in blanks) questions	89	79**
3. Multiple choice and/or true-false questions	81	61**
4. Matching (pairing) questions	69	29**
5. Computational questions	37	32
6. Identification or interpretation questions using maps, graphs, charts, etc.	55	62
B. Oral examinations, class presentations and interviews		
1. Exams which are entirely oral (one student at a time)	19	18
2. Conversations or interviews which are entirely oral (two or more students at a time)	29	34
3. Speeches or other presentations to the class (one student at a time)	44	25**
4. Oral questions combined with written responses (groups of students)	48	54
C. Records of assigned work, classroom participation, major projects, etc.		
1. Records of regularly assigned classwork (e.g. daily problems, laboratory exercises, etc.)	90	72**
2. Records of regularly assigned homework	88	78**
3. Observations of participation in classroom discussion	62	50**
4. Major projects assigned to be completed outside of class (e.g. term papers, research projects, special reports, etc.)	58	31**
D. Performances and demonstrations other than oral		
1. Observations of practical demonstrations using special apparatus, equipment, techniques, etc. (one student at a time)	20	16
2. Observations of practical demonstrations (students working as a team or in a group)	41	33

[a] % indicates the percent of respondents in each sample who employ each technique

** $p < .01$

Table 3

Results of Ratings of Importance of Written Tests in Assigning Grades[1]

		England						Minnesota						
		Ratings				Row		Ratings				Row		
		1	2	3	4	N	%	1	2	3	4	N	%
Subject Matter	E	16	24	26	34	38	20	6	11	36	47	55	26
	FL	3	6	29	62	34	18	0	0	29	71	38	18
	M	4	4	29	64	28	15	2	4	10	84	49	23
	S	5	17	32	46	41	22	3	9	19	69	32	15
	SS	14	14	39	33	49	26	3	16	26	55	38	18
	N	17	26	60	87	190		6	17	52	137	212	
	Col %	9	14	32	46		101	3	8	25	65		101

[1] Cell entries are percentages of respondents by row.

Codes for Subject Matter:
E = English, FL = Foreign Language, M = Mathematics, S = Science, SS = Social Studies

Codes for ratings:
1 = no importance, 2 = minor importance, 3 = moderate importance, 4 = major importance

Results of Hierarchical Log-Linear Analysis

Model	df	χ^2	p
1. (Country, Subject)(Rating)	27	69	<.001
2. (Country, Subject)(Country, Rating)	24	51	<.005
3. (Country, Subject)(Subject, Rating)	15	26	<.05
4. (Country, Subject)(Country, Rating) (Subject, Rating)	12	10	.61

Table 4

Chi Square Values for Hierarchical Log-Linear Analyses of Ratings of Importance of Eleven Factors Considered in Grading[a]

	Model Employed			
Factor Rated	1	2	3	4
End-of-term exam	114	43	93	20*
Oral exam/class presentations	127	121	22**	(15)**
Assigned classwork	47	37	26	15**
Assigned homework	50	36*	24*	7**
Classroom participation	73	57	30	12**
Major projects	154	127	52	13**
Practical demonstrations[b]	43	40	24	20
Class attendance	44	31**	32	(19)*
Student "effort"	119	29**	112	(21)*
Classroom behavior	111	25**	104	(17)**
Work habits	92	32**	71	(11)**

[a] All values are rounded to the nearest integer.

[b] Due to the fact that almost no teachers (3%) rated this factor as "Major", the "Major" category was combined with "Moderate" to create a three-category rating system for purposes of the data analysis.

* $.05 < p < .10$

** $p > .10$

APPENDIX A

Survey Instrument

SURVEY OF GRADING PRACTICES

This survey form is designed to learn about the specific classroom assessment practices employed by teachers in secondary schools. It should require no more than 30 minutes to complete. The focus is upon the procedures used by teachers in evaluating academic achievement in specific courses of study for the purpose of reporting grades to students and their parents. The survey is not concerned with the rating of student effort, attitudes, work habits, etc. where such ratings are separated from grades for subject matter achievement.

The evaluation of student achievement is a difficult and complex process. There are few widely accepted principles to guide a teacher in this task. Therefore, it is not possible to identify "correct" and "incorrect" methods of measuring achievement. Please do not discuss the questions in the survey with your colleagues prior to or while making your responses. Respond to each question by checking the option(s) which best describes your own practices in the course which you designate. Provide a brief description for each question in which your practice is not described by the options given in the survey.

You are not asked to identify yourself. Responses on individual survey forms will be seen only by members of the research staff. Results will be reported by aggregating responses over groups of teachers, i.e. teachers in a single subject area, teachers at a particular grade level, etc. Your willingness to participate in this project is sincerely appreciated.

PLEASE COMPLETE THE FRONT SIDE OF THE SEPARATE RESPONSE SHEET BEFORE RESPONDING TO THE FOLLOWING QUESTIONS.

SECTION 1: EVALUATION TECHNIQUES EMPLOYED

Teachers employ a wide variety of techniques to obtain information on which they base judgments of the academic achievement of students for grade reports. A variety of techniques is listed below under several general headings. Please make a check (✓) on the Response Sheet in the blank in front of each technique you use in the course you have identified.

A. Written tests and examinations

1. essay questions
2. short answer and/or completion (fill in blanks) questions
3. multiple choice and/or true-false questions
4. matching (pairing) questions
5. computational questions
6. identification or interpretation questions using maps, graphs, charts, etc.

B. Oral examinations, class presentations, and interviews

1. exams which are entirely oral (one student at a time)
2. conversations or interviews which are entirely oral (two or more students at a time)
3. speeches or other presentations to the class (one student at a time)
4. oral questions combined with written responses (groups of students)

C. Records of assigned work, classroom participation, major projects, etc.

1. records of regularly assigned class work (e.g. daily problems, laboratory exercises, etc.)
2. records of regularly assigned homework
3. observations of participation in classroom discussions
4. major projects assigned to be completed outside of class (e.g. term papers, research projects, special reports, etc.)

D. Performances and demonstrations other than oral

1. observations of practical demonstrations using special apparatus, equipment, techniques, etc. (one student at a time)
2. observations of practical demonstrations (students working as a team or in a group)

E. Other techniques employed (describe on Response Sheet)

PLEASE CONTINUE TO SECTION 2

SECTION 2: CONSIDERATION GIVEN TO FACTORS IN ASSIGNING GRADES

Teachers typically consider many different types of information and factors when judging achievement for the purpose of assigning subject matter grades at the end of each grading period. A list of such factors is presented below. Indicate how much consideration you give to each factor by selecting one of four response categories: NONE (0%), MINOR (1-10%), MODERATE (11-25%), and MAJOR (26-100%). Please respond on the Response Sheet by writing a "1" for NONE, a "2" for MINOR, a "3" for MODERATE, or a "4" for MAJOR in the blank space corresponding to each factor in the list below.

INFORMATION OR FACTOR

1. Examination given at the end of the marking period
2. Written tests (other than exam given at the end of the marking period)
3. Oral examinations, class presentations, and interviews
4. Regularly assigned class work (problems, exercises, etc.)
5. Regularly assigned homework
6. Participation in classroom discussions
7. Major projects completed outside of class
8. Practical demonstrations
9. Class attendance
10. "Effort" made by the student
11. Student behavior (discipline) in the classroom and school
12. Student work habits (punctuality, neatness, etc.)
13. Other (describe on Response Sheet)
14. Do you use a mathematical formula to weight the above factors when you assign grades? (E.g. 25% tests, 25% assigned class work, 25% homework and 25% class participation. Answer "Yes" or "No" on Response Sheet.)

If you answered "Yes" to question 14, please answer the following question.

15. Do you inform students of the specific weights you use in assigning grades? (Answer "Yes" or "No" on Response Sheet.)

PLEASE CONTINUE TO SECTION 3

SECTION 3: BASIS FOR ASSIGNING GRADES

Indicate which one of the following best represents the primary basis you use in assigning grades to your students by making a check (✓) in the corresponding blank on the Response Sheet.

1. Achievement with respect to individual student ability
2. Achievement with respect to other students of comparable age and grade level
3. Achievement with respect to absolute standards of performance
4. Achievement with respect to individual student improvement or growth
5. Other (describe on Response Sheet)

THANK YOU FOR YOUR COOPERATION